UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/727,886	12/04/2003	Nitendra Rajput	JP920030180US1	8810

7590 10/28/2008

Frederick W. Gibb, III
McGinn & Gibb, PLLC
Suite 304
2568-A Riva Road
Annapolis, MD 21401

EXAMINER	
COLUCCI, MICHAEL C	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
10/28/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/727,886	RAJPUT ET AL.	
	Examiner	Art Unit	
	MICHAEL C. COLUCCI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-21* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-21* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *04 December 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☒ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

1. Applicant's arguments filed 07/29/2008 have been fully considered but they are not persuasive.

The present invention itself in fact teaches the implementation of n-gram models, and more specifically trigram modeling (Present invention page 9) as do both Bahl and Kantrowitz . Further, the present invention also discloses tri-gram modeling with respect to word histories, as is consistent with the teachings of Bahl. The present invention also goes as far to teach a word equivalence probability (Present invention page 13) a two language variant of Bayes theorem. Examiner construes the use of a word equivalence probability to be functionally equivalent and equally effect to the probability that a given word will be next in a series of words in two languages through the use of probabilistic means such as Bayes theorem (Bahl page 1001). Additionally, the secondary reference of Kantrowitz has been incorporated to further strengthen the use of n-gram language models *especially relative to mixed language discourse.*

Even with Bahl teaching tree based analysis, the same goal of next word prediction is taught by Bahl in view of Kantrowitz to address mixed language next word prediction. Just as the present invention, Both Bahl and Kantrowitz both teach statistical/probabilistic approaches to n-gram/trigram language modeling.

Argument 1 (page 10 final paragraph):

- "Bahl merely discloses a standard statistical language model without disclosing how to handle mixed-language data.
- In addition, nowhere does Bahl disclose, teach or suggest storing word equivalence probabilities. Instead, Bahl describes partitioning word histories into equivalence classes, such as, N-grams, without mentioning mixed-language data. Bahl's disclosure relates to a tree-based statistical language model for natural language speech recognition in but one language."

Response to argument 1 :

Bahl clearly teaches within the abstract, a remedy to the problem of <u>"predicting" the next word a speaker will say, given the words already spoken; specifically, the problem is to estimate the probability that a given word will be the next word uttered. Algorithms are presented for automatically constructing a binary decision tree designed to estimate these probabilities</u> (Abstract).

Examiner takes the position that Bahl in fact teaches storing and the use of word equivalence probabilities, wherein Bahl teaches a probabilistic approach to determine the next spoken word based on previous words, where a language model equivalence classes is used. Bahl teaches that all words prior to the most recent two are ignored, and useful information is lost. Additionally, word sequences ending in different pairs of words should not necessarily be

considered distinct; they may be functionally equivalent from a language model point of view. Separating equivalent histories into different classes, as the trigram model does, fragments the training data unnecessarily and reduces the accuracy of the resulting probability estimates. Further, Bahl teaches that at each nonterminal node of the tree, there is a question requiring a yes/no answer, and corresponding to each possible answer there is a branch leading to the next question. Associated with each terminal node, i.e., leaf, is some advice or information which takes into account all the questions and answers which lead to that leaf. In the context of language modeling, the questions relate to the words already spoken; for example: "Is the preceding word a verb?". And the information at each leaf takes the form of a probability distribution indicating which words are likely to be spoken next. The leaves of the tree represent language model equivalence classes. (page 1002 Col. 1 & Fig. 1).

With Bahl teaching a single language modeling and probabilistic word prediction, Kantrowitz is incorporated to address a mixed language, wherein Kantrowitz certainly teaches the handling of a mixed language document in a statistical manner. Like the present invention (Present invention page 7),Kantrowitz teaches a word by word basis approach to statistical analysis. Kantrowitz teaches a method that is different from these systems in that it identifies the language of individual words with very high accuracy, not entire documents. This allows the present invention to operate on a word-by-word basis, correctly

identifying the language of words even when the document contains multiple languages (e.g., Canadian parliamentary proceedings contain both English and French) or includes short quotes of one language within a document that is mostly another language. This allows language-specific functionality, such as language-specific spelling correction and transliteration (e.g., ASCII-to-Kanji conversion of Japanese Romaji to Kanji letters) to occur on a word-by-word basis. The language identification statistics for the individual words of a document can be combined to identify the overall language of a document with much higher cumulative accuracy than the state of the art. It can also identify the number of languages present in mixed-language documents, the identity of the language and the relative frequency of occurrence of the language's lexicon (Kantrowitz Col. 2 lines 17-47).

Further, Kantrowitz does not just merely disclose the identification of words in a mixed language document in a standard manner. Kantrowitz teaches the elimination of burdensome user intervention allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. In computer user interfaces, it can be used to automatically select the language in which the

system interacts with the user (e.g., the language of menus and help systems), to identify the source language for machine translation applications without requiring the user to explicitly specify the source language, and to identify the most likely ancestry and/or native language of a person by identifying the language of their name (Kantrowitz Col. 6 lines 7-26).

Kantrowitz teaches that the invention herein goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another (Kantrowitz Col. 6 lines 41-67). The missing element from the scope of the invention is performing the methods taught by Bahl in a mixed language environment, and thus Kantrowitz is

introduced to take the teachings of Bahl to the next level performing probabilistic approaches to word prediction in a mixed language environment.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate using word histories and probabilities for statistical purposes using parallel identifiers for specific languages relative to a lexicon/corpus, where the next word in a mixed language text can be predicted as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text.

Argument 2 (page 11 paragraphs 2 and 3, page 12 paragraphs 2 and 3):

- "Instead, Bahl merely discloses a standard statistical language model without disclosing how to handle mixed-language data. Kantrowitz merely discloses identifying individual words of mixed languages in a document"

AND

- "Instead, Kantrowitz merely discloses identifying individual words of mixed languages in a document"

Response to argument 2 :

Bahl clearly teaches within the abstract, a remedy to the problem of "predicting" the next word a speaker will say, given the words already spoken; specifically, the problem is to estimate the probability that a given word will be the next word uttered. Algorithms are presented for automatically constructing a binary decision tree designed to estimate these probabilities (Abstract).

Examiner takes the position that Bahl in fact teaches storing and the use of word equivalence probabilities, wherein Bahl teaches a probabilistic approach to determine the next spoken word based on previous words, where a language model equivalence classes is used. Bahl teaches that all words prior to the most recent two are ignored, and useful information is lost. Additionally, word sequences ending in different pairs of words should not necessarily be considered distinct; they may be functionally equivalent from a language model point of view. Separating equivalent histories into different classes, as the trigram model does, fragments the training data unnecessarily and reduces the accuracy of the resulting probability estimates. Further, Bahl teaches that at each nonterminal node of the tree, there is a question requiring a yes/no answer, and corresponding to each possible answer there is a branch leading to the next

question. Associated with each terminal node, i.e., leaf, is some advice or information which takes into account all the questions and answers which lead to that leaf. In the context of language modeling, the questions relate to the words already spoken; for example: "Is the preceding word a verb?". And the information at each leaf takes the form of a probability distribution indicating which words are likely to be spoken next. The leaves of the tree represent language model equivalence classes. (page 1002 Col. 1 & Fig. 1).

With Bahl teaching a single language modeling and probabilistic word prediction, Kantrowitz is incorporated to address a mixed language, wherein Kantrowitz certainly teaches the handling of a mixed language document in a statistical manner. Like the present invention (Present invention page 7),Kantrowitz teaches a word by word basis approach to statistical analysis. Kantrowitz teaches a method that is different from these systems in that it identifies the language of individual words with very high accuracy, not entire documents. This allows the present invention to operate on a word-by-word basis, correctly identifying the language of words even when the document contains multiple languages (e.g., Canadian parliamentary proceedings contain both English and French) or includes short quotes of one language within a document that is mostly another language. This allows language-specific functionality, such as language-specific spelling correction and transliteration (e.g., ASCII-to-Kanji conversion of Japanese Romaji to Kanji letters) to occur on a word-by-word

basis. The language identification statistics for the individual words of a document can be combined to identify the overall language of a document with much higher cumulative accuracy than the state of the art. It can also identify the number of languages present in mixed-language documents, the identity of the language and the relative frequency of occurrence of the language's lexicon (Kantrowitz Col. 2 lines 17-47).

Further, Kantrowitz does not just merely disclose the identification of words in a mixed language document in a standard manner. Kantrowitz teaches the elimination of burdensome user intervention allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. In computer user interfaces, it can be used to automatically select the language in which the system interacts with the user (e.g., the language of menus and help systems), to identify the source language for machine translation applications without requiring the user to explicitly specify the source language, and to identify the most likely ancestry and/or native language of a person by identifying the language of their name (Kantrowitz Col. 6 lines 7-26).

Kantrowitz teaches that the invention herein goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another (Kantrowitz Col. 6 lines 41-67). The missing element from the scope of the invention is performing the methods taught by Bahl in a mixed language environment, and thus Kantrowitz is introduced to take the teachings of Bahl to the next level performing probabilistic approaches to word prediction in a mixed language environment.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate using word histories and probabilities for statistical purposes using parallel identifiers for specific languages relative to a lexicon/corpus, where the next word in a mixed

language text can be predicted as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-21 rejected under 35 U.S.C. 103(a) as being unpatentable over Bahl et al., "A tree-based statistical language model for natural language speech recognition" (hereinafter Bahl) in view of Kantrowitz US 6292772 B1 (hereinafter Kantrowitz).

Re claims 1, 8, and 9, Bahl teaches storing word equivalence probabilities relating to words of a first language and words in at least one other language (Page 1001 Col. 2);

generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence (Page 1001 Col. 2);

generating monolingual next word hypothesis probabilities (Page 1002 Col. 2) in the first language based upon the monolingual word history (Page 1001 Col. 2), wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence (Page 1006 Col. 1 paragraphs 1-3);

determining a probability of a next word (Page 1002 Col. 2) in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities (Page 1001 Col. 2), wherein said probability of said next word predicts a next word in said mixed language expression (Page 1006 Col. 1 paragraphs 1-3)

However, Bahl fails to teach a method for language modeling of mixed language expressions (Kantrowitz Col. 6 lines 7-64)

Bahl teaches that all current Japanese word processing systems require the user to explicitly switch from a Japanese mode into an English mode. The same is true of

other foreign language word processing systems, where the user must explicitly state the target language. The present invention eliminates this step, allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. Kantrowitz also teaches the method of recognizing the language of a single word has applications to spelling and grammar correction (e.g., identifying the appropriate language resources on a document, paragraph, sentence or even individual word basis), the automatic invocation of transliteration software based on the language of the words (e.g., automatic ASCII to Kanji substitution without requiring the user to explicitly switch into a Kanji mode), the automatic invocation of appropriate machine translation tools when the document's language is different from the user's native tongue(s), the use of document language identification to eliminate from database or web search results any documents which are not written in the user's native languages and the automatic identification of user-appropriate languages for the user interface.

Additionally, Kantrowitz teaches that the present invention determines whether or not a word is in the lexicon of a specific language. It is possible that a word may satisfy the recognizer (statement of n-gram patterns) for more than one language, using multiple parallel recognizers for specific languages, we can identify the languages to which the word belongs. If a word matches several recognizers, one can either weigh

each language equally or use the language of the words on the left and right to disambiguate the possibilities. For example, if both neighboring words are English and the current word is recognized as being both English and Japanese, the current word would be deemed to be English. On the other hand, if one of the neighboring words was Japanese, both English and Japanese would be reported.

Further, Kantrowitz teaches that the invention goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another. This invention is not limited to comparing only two languages. First, a collection of regular expressions for pair wise distinguishing languages can be used to identify the language of a word. Moreover, lexicons of multiple languages could be merged to distinguish, for example, English words from the words present in any one of a dozen other languages.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate using word histories and probabilities for statistical purposes using parallel identifiers for specific languages relative to a lexicon/corpus, where the next word in a mixed language text can be predicted as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text (Kantrowitz Col. 6 lines 7-26).

Re claims 2, 10, and 16, Bahl teaches the method as claimed in claim 1, further comprising summing products of word equivalence probabilities with respective monolingual next word hypothesis probabilities (Page 1002 Col. 2).

Re claims 3, 11, and 17, Bahl teaches the method as claimed in claim 1, wherein the monolingual next word (Page 1002 Col. 2) hypothesis probability is a statistical language model (Page 1001 Col. 1).

Re claims 4, 12, and 18, Bahl fails to teach the method as claimed in claim 1, further comprising converting a mixed language word sequence to a monolingual word sequence using word equivalence probabilities (Kantrowitz Col. 6 lines 7-64).

Bahl teaches that all current Japanese word processing systems require the user to explicitly switch from a Japanese mode into an English mode. The same is true of other foreign language word processing systems, where the user must explicitly state the target language. The present invention eliminates this step, allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. Kantrowitz also teaches the method of recognizing the language of a single word has applications to spelling and grammar correction (e.g., identifying the appropriate language resources on a document, paragraph, sentence or even individual word basis), the automatic invocation of transliteration software based on the language of the words (e.g., automatic ASCII to Kanji substitution without requiring the user to explicitly switch into a Kanji mode), the automatic invocation of appropriate machine translation tools when the document's language is different from the user's native tongue(s), the use of document language identification to eliminate from database or web search results any documents which are not written in the user's native languages and the automatic identification of user-appropriate languages for the user interface.

Additionally, Kantrowitz teaches that the present invention determines whether or not a word is in the lexicon of a specific language. It is possible that a word may satisfy the recognizer (statement of n-gram patterns) for more than one language, using multiple parallel recognizers for specific languages, we can identify the languages to which the word belongs. If a word matches several recognizers, one can either weigh each language equally or use the language of the words on the left and right to disambiguate the possibilities. For example, if both neighboring words are English and the current word is recognized as being both English and Japanese, the current word would be deemed to be English. On the other hand, if one of the neighboring words was Japanese, both English and Japanese would be reported.

Further, Kantrowitz teaches that the invention goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another. This invention is not limited to comparing only two languages.

First, a collection of regular expressions for pair wise distinguishing languages can be used to identify the language of a word. Moreover, lexicons of multiple languages could be merged to distinguish, for example, English words from the words present in any one of a dozen other languages.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate converting a mixed language word sequence to a monolingual word sequence using word equivalence probabilities as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text (Kantrowitz Col. 6 lines 7-26).

Re claims 5, 13, and 19, Bahl teaches the method as claimed in claim 1, further comprising determining the word equivalence probabilities (Page 1001 Col. 2).

However, Bahl fails to teach a parallel text corpus that has corresponding expressions in the first language and the at least one other language (Kantrowitz Col. 6 lines 7-64).

Kantrowitz teaches that all current Japanese word processing systems require the user to explicitly switch from a Japanese mode into an English mode. The same is true of other foreign language word processing systems, where the user must explicitly state the target language. The present invention eliminates this step, allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. Kantrowitz also teaches the method of recognizing the language of a single word has applications to spelling and grammar correction (e.g., identifying the appropriate language resources on a document, paragraph, sentence or even individual word basis), the automatic invocation of transliteration software based on the language of the words (e.g., automatic ASCII to Kanji substitution without requiring the user to explicitly switch into a Kanji mode), the automatic invocation of appropriate machine translation tools when the document's language is different from the user's native tongue(s), the use of document language identification to eliminate from database or web search results any documents which are not written in the user's native languages and the automatic identification of user-appropriate languages for the user interface.

Additionally, Kantrowitz teaches that the present invention determines whether or not a word is in the lexicon of a specific language. It is possible that a word may satisfy the recognizer (statement of n-gram patterns) for more than one language, using

multiple parallel recognizers for specific languages, we can identify the languages to which the word belongs. If a word matches several recognizers, one can either weigh each language equally or use the language of the words on the left and right to disambiguate the possibilities. For example, if both neighboring words are English and the current word is recognized as being both English and Japanese, the current word would be deemed to be English. On the other hand, if one of the neighboring words was Japanese, both English and Japanese would be reported.

Further, Kantrowitz teaches that the invention goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another. This invention is not limited to comparing only two languages. First, a collection of regular expressions for pair wise distinguishing languages can be used to identify the language of a word. Moreover, lexicons of multiple languages could

be merged to distinguish, for example, English words from the words present in any one of a dozen other languages.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate a parallel text corpus that has corresponding expressions in the first language and the at least one other language as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text (Kantrowitz Col. 6 lines 7-26).

Re claims 6, 14, and 20, Bahl teaches the method as claimed in claim 1, further comprising determining a probability of a next word (Page 1002 Col. 2) hypothesis given a base language word history (Page 1001 Col. 2).

However, Bahl fails to teach probabilities of a foreign language given a base language (Kantrowitz Col. 6 lines 7-64).

Kantrowitz teaches that all current Japanese word processing systems require the user to explicitly switch from a Japanese mode into an English mode. The same is

true of other foreign language word processing systems, where the user must explicitly state the target language. The present invention eliminates this step, allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. Kantrowitz also teaches the method of recognizing the language of a single word has applications to spelling and grammar correction (e.g., identifying the appropriate language resources on a document, paragraph, sentence or even individual word basis), the automatic invocation of transliteration software based on the language of the words (e.g., automatic ASCII to Kanji substitution without requiring the user to explicitly switch into a Kanji mode), the automatic invocation of appropriate machine translation tools when the document's language is different from the user's native tongue(s), the use of document language identification to eliminate from database or web search results any documents which are not written in the user's native languages and the automatic identification of user-appropriate languages for the user interface.

Additionally, Kantrowitz teaches that the present invention determines whether or not a word is in the lexicon of a specific language. It is possible that a word may satisfy the recognizer (statement of n-gram patterns) for more than one language, using multiple parallel recognizers for specific languages, we can identify the languages to which the word belongs. If a word matches several recognizers, one can either weigh

each language equally or use the language of the words on the left and right to disambiguate the possibilities. For example, if both neighboring words are English and the current word is recognized as being both English and Japanese, the current word would be deemed to be English. On the other hand, if one of the neighboring words was Japanese, both English and Japanese would be reported.

Further, Kantrowitz teaches that the invention goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another. This invention is not limited to comparing only two languages. First, a collection of regular expressions for pair wise distinguishing languages can be used to identify the language of a word. Moreover, lexicons of multiple languages could be merged to distinguish, for example, English words from the words present in any one of a dozen other languages.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate probabilities of a foreign language given a base language as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text (Kantrowitz Col. 6 lines 7-26).

Re claims 7, 15, and 21, Bahl fails to teach the method as claimed in claim 1, further comprising using a parallel text corpus that has corresponding expressions in the first language and the at least one other language (Kantrowitz Col. 6 lines 7-64)

Kantrowitz teaches that all current Japanese word processing systems require the user to explicitly switch from a Japanese mode into an English mode. The same is true of other foreign language word processing systems, where the user must explicitly state the target language. The present invention eliminates this step, allowing the user to type in English or Romaji as needed, with the system automatically distinguishing between the two and converting the Romaji to Kanji as necessary. In a mixed-language document, this regular expression can be used to select the appropriate dictionary and

thesaurus for use with the word. It can also be used to select the appropriate spelling correction and grammar correction algorithms. Kantrowitz also teaches the method of recognizing the language of a single word has applications to spelling and grammar correction (e.g., identifying the appropriate language resources on a document, paragraph, sentence or even individual word basis), the automatic invocation of transliteration software based on the language of the words (e.g., automatic ASCII to Kanji substitution without requiring the user to explicitly switch into a Kanji mode), the automatic invocation of appropriate machine translation tools when the document's language is different from the user's native tongue(s), the use of document language identification to eliminate from database or web search results any documents which are not written in the user's native languages and the automatic identification of user-appropriate languages for the user interface.

Additionally, Kantrowitz teaches that the present invention determines whether or not a word is in the lexicon of a specific language. It is possible that a word may satisfy the recognizer (statement of n-gram patterns) for more than one language, using multiple parallel recognizers for specific languages, we can identify the languages to which the word belongs. If a word matches several recognizers, one can either weigh each language equally or use the language of the words on the left and right to disambiguate the possibilities. For example, if both neighboring words are English and the current word is recognized as being both English and Japanese, the current word would be deemed to be English. On the other hand, if one of the neighboring words was Japanese, both English and Japanese would be reported.

Further, Kantrowitz teaches that the invention goes beyond the state of the art by being able to identify the language of individual words in isolation with high accuracy. The accuracy in identifying the language of individual words typically is equal to that of whole-document language identification systems. When the language identification of individual words is combined for all the words in a document, the overall accuracy significantly exceeds that of whole-document systems. Moreover, the ability to identify the language of individual words permits document processing resources to be applied on a word-by-word basis. For example, it allows for the spelling correction of a mixed-language document, allowing the spelling correction software to select the appropriate language for each word. It also allows the automatic substitution of Kanji for Romaji in mixed Japanese-English documents, without requiring the user to explicitly switch from one language to another. This invention is not limited to comparing only two languages. First, a collection of regular expressions for pair wise distinguishing languages can be used to identify the language of a word. Moreover, lexicons of multiple languages could be merged to distinguish, for example, English words from the words present in any one of a dozen other languages.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Bahl to incorporate using a parallel text corpus that has corresponding expressions in the first language and the at least one other language as taught by Kantrowitz because using word prediction and probabilities relative to a mixed language allows for an interface that enables a multilingual user to input a language that may have two or more mixed languages, wherein the ability to

model a mixed language text allows for multiple languages in one text to be distinguished from one another, without translation from one language to another, where automatic substitution of words occurs through the use of various lexicons and additionally, lexicons may be merged to allow for an increased capability of modeling additional languages mixtures for the purpose of predicting a more versatile selection of adjacent words in a text (Kantrowitz Col. 6 lines 7-26).

Conclusion

4. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-

270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner
AU 2626
(571)-270-1847
Michael.Colucci@uspto.gov

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626